**ARE SUICIDE RATES LINKED TO THEIR COUNTRY’S ECONOMIC GROWTH?**

It sounds like a no brainer to assume that suicide rate is majorly linked to economic growth/fall of a country. But is that true? Let us find out by analyzing the suicide data taken from Kaggle. The goal of this analysis is to find out if there is any relationship between Suicide rates and GDP per capita.

**Data and metrics used for the study:**

The data for this study is a csv file that includes suicide information taken from 101 countries over a span of three decades and the fields are:

country, year, sex, age group, count of suicides, population, suicide rate, country-year composite key, HDI for year, gdp\_for\_year, gdp\_per\_capita, generation (based on age grouping average).

We have 3 fields, HDI for year, gdp for year and gdp per capita, in the dataset that give information about the economic situation of a country. The Human Development Index (HDI) is an index that measures key dimensions of human development and it would be a good fit for our study. But taking a closer look at our data, HDI is only available for recent years. Gross Domestic Product (GDP) for year is a monetary measure of the market value of all the final goods and services produced in an year. Since it describes the overall economic conditions of a country as a whole, I have opted for GDP per capita as it is often used as an indicator of living standards.

**Challenges with this dataset:**

For this analysis, I had to clean the data to remove unwanted columns, normalize suicides count by extracting number of suicides per 100K population, aggregate suicides and population based of country and year. HDI and GDP for year information is removed as GDP per capita would be a better metric to determine the standard of living of an average person in a given country.

--Histogram with mean GDP per capita vs mean suicide numbers for 76 countries

**Looking at the initial readings:**

I have taken a correlation between GDP per capita and number of suicides per country on the key dataset for the above 76 countries.

Count of countries that have high correlation from 76 countries: 50

Number of countries with a positive correlation: 14

Number of countries with a negative correlation: 36

**How trends are varying over countries from different economic backgrounds:**

I have categorized the correlation data into 3 crucial sets that would help me analyze the correlation efficiently. The three sets are:

1. Countries with high GDP per capita where both max and min GDP per capita are high

There are 10 countries in total for the above category. Out of those 10 countries, 2 have high positive correlation and 4 have high negative correlation.

There are 5 countries with low correlation

2. Countries with low GDP per capita where both max and min GDP per capita and low

3. Countries that had low GDP per capita for min GDP and high GDP per capita for max GDP

**My assumptions:**

**Conclusion:**

**Future work:**

-The initial step was to group the data first by country and then by year. Aggregation was done as followed:

Applied Sum to Population and Suicides\_no. Applied Max to gdp\_per\_capita\_in\_dollars. Also, a new column ‘suicides\_per\_100k\_people’ was added to making the readings consistent. This is the key dataset that was used to find out the correlation between GDP per capita and number of suicides

- Second step was to calculate the following fields:

Country, count\_of\_years, mean\_gdp\_per\_capita\_in\_dollars, gdp\_min, gdp\_max, gdp\_diff, gdp\_timesgrowth

count\_of\_years: count of number of years of data per country

mean\_gdp\_per\_capita\_in\_dollars: Mean of GDP per capita for all the years per country

gdp\_min, gdp\_max: Minimum and maximum GDP per capita values from the given years per country

gdp\_timesgrowth: Variation between the minimum and maximum GDP for the given years per country

- Since I was trying to perform analysis on a reasonable number of data points, I have filtered out the countries that do not have at least 20 years of information. This filter has reduced the dataset to 76 countries out of 101 countries

- I have taken a correlation between GDP per capita and number of suicides per country on the key dataset for the above 76 countries.

Count of countries that have high correlation from 76 countries: 50

Number of countries with a positive correlation: 14

Number of countries with a negative correlation: 36

- I have categorized the correlation data into 3 crucial sets that would help me analyze the correlation efficiently. The three sets are:

1. Countries with high GDP per capita where both max and min GDP per capita are high

There are 10 countries in total for the above category. Out of those 10 countries, 2 have high positive correlation and 4 have high negative correlation.

There are 5 countries with low correlation

2. Countries with low GDP per capita where both max and min GDP per capita and low

3. Countries that had low GDP per capita for min GDP and high GDP per capita for max GDP

Notes: 1. Criteria for GDP scale, and correlation (above 0.5 and below -0.5)

2. Conclude the analysis

3. Scatter plots with countries

4. Read the question in thinkful one more time

5. Find out if the stats abt the dataset. No of countries, years range, gdp range, how many countries with high GDP, how many countries with low GDP.